



## PATENT ABSTRACTS OF JAPAN

(11) Publication number: **05165494 A**(43) Date of publication of application: **02.07.93**

(51) Int. Cl. **G10L 3/00**  
**G10L 3/00**  
**G10L 3/00**

(21) Application number: **03330807**(22) Date of filing: **13.12.91**(71) Applicant: **OSAKA GAS CO LTD OKI  
ELECTRIC IND CO LTD**

(72) Inventor: **OBA KATSUYA  
HIRAYAMA TERU  
TAGAWA TADAMICHI  
KATO MASAOKI**

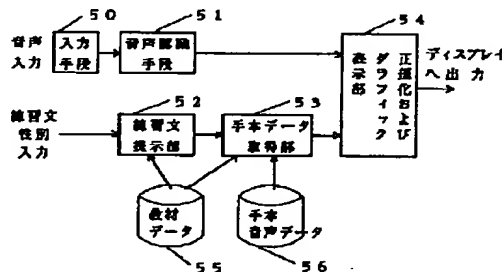
**(54) VOICE RECOGNIZING DEVICE****(57) Abstract:**

**PURPOSE:** To obtain a low-priced system facilitating the study of a learning person without being hourly limited by providing a correspondence display means to display the prescribed acoustic feature amount of a voice signal and a language symbol corresponding to the amount while making them correspondent.

**CONSTITUTION:** This device is equipped with an input means 50 to input the voice of an unspecified speaker, voice recognizing means 51 to obtain the language symbol by executing voice recognition according to the voice signal, and display means 54 to display the contents of making the prescribed acoustic characteristic amount of the voice signal correspondent to the language symbol corresponding to the amount by normalizing them concerning the two kinds of voices at least while making those amount and symbol correspondent. Namely, a normalizing part and graphic display part 54 inputs the voice data of the learning person outputted from the voice recognizing means 51 and model voice data from a model data acquisition part 52, normalizes speaking time for each word to which both data are correspondent, and graphically displays the

voice data of the learning person and the model voice data (intonation, stress).

COPYRIGHT: (C)1993,JPO&amp;Japio



(19)日本国特許庁(JP)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開平5-165494

(43)公開日 平成5年(1993)7月2日

(51)Int.Cl. <sup>5</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 1 0 L 3/00	5 5 1 E	8842-5H		
	S	8946-5H		
	5 6 1 C	8842-5H		

審査請求 未請求 請求項の数5(全 18 頁)

(21)出願番号 特願平3-330807

(22)出願日 平成3年(1991)12月13日

(71)出願人 000000284

大阪瓦斯株式会社

大阪府大阪市中央区平野町四丁目1番2号

(71)出願人 000000295

沖電気工業株式会社

東京都港区虎ノ門1丁目7番12号

(72)発明者 大場 克哉

大阪府豊中市新千里西町1丁目2番1号

株式会社オージー情報システム総研内

(72)発明者 平山 輝

大阪市中央区平野町四丁目1番2号 大阪

瓦斯株式会社内

(74)代理人 弁理士 松田 正道

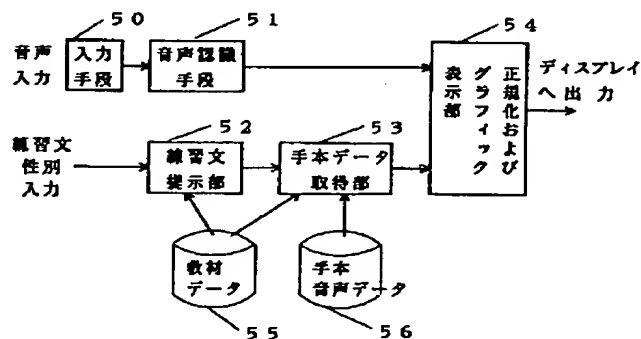
最終頁に続く

(54)【発明の名称】 音声認識装置

(57)【要約】

【目的】時間的制約を受けず、安価で、しかも学習者の学習が容易であるようなシステムを実現できる音声認識装置を提供すること。

【構成】不特定話者の音声を入力する入力手段50と、その音声信号から音声認識を行い、言語シンボルを得る認識手段51と、前記音声信号の所定の音響的特徴量とそれに対応する前記言語シンボルとの対応付け内容を、少なくとも2種類の音声について、正規化して対応させ表示する表示手段54とを備えた音声認識装置である。



## 【特許請求の範囲】

【請求項1】 音声を入力する入力手段と、その音声信号から音声認識を行い、言語シンボルを得る認識手段と、前記音声信号の所定の音響的特徴量とそれに対応する前記言語シンボルとを対応付けながら表示する対応表示手段とを備えたことを特徴とする音声認識装置。

【請求項2】 不特定話者の音声を入力する入力手段と、その音声信号から音声認識を行い、言語シンボルを得る認識手段と、前記音声信号の所定の音響的特徴量とそれに対応する前記言語シンボルとの対応付け内容を、少なくとも2種類の音声について、正規化して対応させ表示する表示手段とを備えたことを特徴とする音声認識装置。

【請求項3】 不特定話者の音声を入力する入力手段と、その音声信号から音声認識を行い、言語シンボルを得る認識手段と、前記音声信号の所定の音響的特徴量とそれに対応する前記言語シンボルとの対応付け内容を、少なくとも2種類の音声について、比較する比較手段とを備えたことを特徴とする音声認識装置。

【請求項4】 前記2種類の音声は、手本となる音声と、学習者の音声であり、教育目的に用いられることを特徴とする請求項2又は3記載の音声認識装置。

【請求項5】 前記手本となる音声についての前記対応付け内容は、予め獲得され、記憶手段に格納されていることを特徴とする請求項4記載の音声認識装置。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】 本発明は、教育などの分野に応用可能な音声認識装置に関するものである。

## 【0002】

【従来の技術】 従来、例えば、外国語教育における発音（スピーキング）の訓練を行う方法としては次のようなものが知られている。（1）外国語教育専門家やネイティブスピーカーに自分の発音を聞いてもらい、アドバイスを受け、学習者の欠点を指摘してもらう。（2）市販の外国語学習用テープや2カ国語放送などを聞き、それをまねて発音する。また、参考書などに記載されている口の形や調音点等を示す図表をもとに練習する。自分の発音を、磁気テープなどに録音し、学習用テープに録音されている音声と聞き比べてみる。（3）計算機等を利用した発音練習装置により練習する。すなわち、発音練習用のCAI（computer assisted instruction：コンピュータ利用の教育）システムが知られている。これは、ネイティブスピーカーの発音を聞く。学習者の声を録音し、再生する。ネイティブスピーカーの音声波形と、学習者の音声波形をグラフィック表示する。また、ホルマントの表示や、時間毎の周波数や音圧の変化が表示される。

## 【0003】

【発明が解決しようとする課題】 しかしながら、上記の

ような従来の発音練習方法では次のような課題がある。

（1）外国語教育専門家やネイティブスピーカーによる訓練方法では、大きな効果が期待できるが、時間的な制約を受け、また費用もかかる。（2）音声テープなどによる方法では、いつでも練習でき、また安価ではあるが、発した発音を客観的に評価できないため、学習者は自分の発音の何処が良くないかを把握することが困難である。（3）計算機などを利用して発音練習を行う方法では、確かにネイティブスピーカーの音声波形と学習者の音声波形が表示されるが、そこから何をどの様に読み取れば、学習者の発音がネイティブスピーカーの発音に近づくかが分かりにくいという課題がある。

【0004】 本発明は、このような従来の発音訓練方法の課題を考慮し、時間的制約を受けず、安価で、しかも学習者の学習が容易であるようなシステムを実現できる音声認識装置を提供することを目的とするものである。

## 【0005】

【課題を解決するための手段】 第1の本発明は、音声を入力する入力手段と、その音声信号から音声認識を行い、言語シンボルを得る認識手段と、前記音声信号の所定の音響的特徴量とそれに対応する前記言語シンボルとを対応付けながら表示する対応表示手段とを備えた音声認識装置である。

【0006】 第2の本発明は、不特定話者の音声を入力する入力手段と、その音声信号から音声認識を行い、言語シンボルを得る認識手段と、前記音声信号の所定の音響的特徴量とそれに対応する前記言語シンボルとの対応付け内容を、少なくとも2種類の音声について、正規化して対応させ表示する表示手段とを備えた音声認識装置である。

【0007】 第3の本発明は、不特定話者の音声を入力する入力手段と、その音声信号から音声認識を行い、言語シンボルを得る認識手段と、前記音声信号の所定の音響的特徴量とそれに対応する前記言語シンボルとの対応付け内容を、少なくとも2種類の音声について、比較する比較手段とを備えた音声認識装置である。

## 【0008】

【作用】 第1の本発明では、入力手段で入力された音声の信号から、認識手段によって音声認識を行い、言語シンボルを得る。また、対応表示手段は、音声信号の所定の音響的特徴量とそれに対応する言語シンボルとを対応付けながら表示する。

【0009】 第2の本発明では、入力手段で入力された不特定話者の音声信号から音声認識を行い、言語シンボルを得る。表示手段は、音声信号の所定の音響的特徴量とそれに対応する言語シンボルとの対応付け内容を、少なくとも2種類の音声について、正規化して表示する。

【0010】 第2の本発明では、入力手段で入力された不特定話者の音声信号から音声認識を行い、言語シンボルを得る。音声信号の所定の音響的特徴量とそれに対応

する言語シンボルとの対応付け内容を、比較手段によって、少なくとも2種類の音声について、比較する。

【0011】

【実施例】以下本発明の一実施例について図面を参照しながら説明する。

【0012】図1は本発明の音声認識装置の一実施例である音声データのグラフィック表示を行うための、コンピュータを利用したシステムを示す図である。同図に於て、学習者の音声マイク15から入力され、不特定話者連続音声認識手段14により解析される。本発明の処理を行うプログラムと必要なデータが磁気ディスク13に格納されており、CPU11により、メモリ12にロードされる。キーボード17、マウス18は、後に述べる練習文を選択する際に使用され、イントネーション、ストレスなどがディスプレイ16にグラフィック表示されるようになっている。

【0013】図2は、本発明で使用する不特定話者連続音声認識手段の出力の一例である音声データである。認識した単語数21、各単語の文字列の配列22、各単語のセグメント（認識された音素）数の配列23、各セグメントのピッチ周波数の平均の配列24、各セグメントの音圧の平均の配列25、各セグメントの発話時間の配列26からなる。

【0014】以下の実施例1は、第1の本発明の一実施例であって、認識対象である音声信号のイントネーションまたはストレスを、認識結果である単語に対応させて表示する装置を示す。

【0015】また、以下の説明で配列のn番目とは、最初の要素を0番目とし、以下1番目、2番目、...とする。

【0016】（実施例1）図3は、本発明の実施例1のブロック図である。マイクロフォン15等の入力手段30から学習者の音声を入力し、図2に示した音声データを出力する、認識手段の一例としての音声認識処理部31、および音声認識処理部31から出力される音声データを入力とし、これから入力音声のイントネーション、ストレスを認識結果の単語列に対応させてディスプレイに表示する、対応表示手段の一例としてのグラフィック表示部32からなる。

【0017】図4は、その音声認識処理部31の内容を示すフローチャートである。すなわち図4で、ステップ310では、マイクロフォン15からの音声信号についてアナログ処理、A/D変換がおこなわれる。ここで、入力：アナログ連続音声信号、出力：デジタル音声信号、処理概要：アンチ・エイリアスフィルタ、サンプリング周波数16kHzで、デジタル信号に変換するものである。

【0018】ステップ311では、音響パラメータ変換がおこなわれる。ここで、入力：デジタル音声信号、出力：フレーム（6.6ミリ秒）毎の23の音響的特徴

量、処理概要：デジタル信号を6.6ミリ秒のフレームに分け、スペクトル分析などにより、23の音響的特徴量を抽出するものである。

【0019】ステップ312では、音素コード（音素片）変換がおこなわれる。ここで、入力：フレーム毎の音響的特徴量、出力：セグメント毎の音素コード（音素片）、およびセグメント毎の音響的特徴量、処理概要：フレーム毎の音響的特徴量をもとに各フレームに音素コード（約1800種類）を割り当てるものである。似通った特徴を持つ隣あったフレームはまとめて、1つのセグメントとするものである。

【0020】ステップ313では、音素コード（音素片）／音素変換がおこなわれる。ここで、入力：セグメント毎の音素コード（音素片）およびセグメント毎の音響的特徴量、出力：セグメント毎の音素の候補、およびセグメント毎の音響的特徴量、外部参照データ：音素コードブック315、処理概要：音素コードブックを参照し、セグメント毎にそのセグメントが各音素（約50種類）に対応する可能性を割り当てる。

【0021】ステップ314では、音素／単語列変換がおこなわれる。ここで、入力：セグメント毎の音素の候補およびセグメント毎の音響的特徴量、出力：単語列およびセグメント毎の音響的特徴量、外部参照データ：単語辞書316、処理概要：単語辞書を参照し、入力のセグメント列と各単語に対応する音素列との距離を計算し、最も近いものを認識結果として出力する。

【0022】なお、上記音素コードブック315は、音素コード（音素片）毎にその音素コードがある音素である可能性をもつテーブルであり、また、上記単語辞書316は、単語文字列とそれに対応する音素列を持つ辞書である。

【0023】図5及び図6は、図3のグラフィック表示部32の処理フローの例である。

【0024】401は、前処理であり、ディスプレイにX軸、Y軸を表示し、また、各変数の初期化を行う。すなわち、処理中の単語が認識結果の単語列の何番目かを示す変数L\_word\_cnt、処理中のセグメントが認識結果のセグメントの配列の何番目かを示す変数L\_seg\_cnt、グラフィック表示する折れ線グラフ（後述する図15参照）の各標本点の値を示す変数（配列）の0番目L\_x[0]、L\_y[0]にそれぞれ0を代入する。

【0025】402では、音声認識結果の音声データの単語数をL\_word\_numに代入する。

【0026】403では、L\_word\_cntとL\_word\_numの比較を行う。L\_word\_cntがL\_word\_numより小さければ、処理すべき単語が存在するという事であり、ステップ404へ進む。L\_word\_cntがL\_word\_num以上であれば全ての単語についてグラフィック表示が終了したということであり、全ての処理を終える。

【0027】404では、次の折れ線グラフ表示時にX

軸の下部に現在処理している単語を表示することを示すフラグdisp\_word\_flagをONにする。

【0028】405では、現在処理している単語が何番目のセグメントまでかを示す変数L\_seg\_numに音声データの各単語のセグメント数の配列のL\_word\_cnt番目の要素を加える。

【0029】406では、L\_seg\_cntとL\_seg\_numとの比較を行う。L\_seg\_cntがL\_seg\_numより小さければ、現在処理している単語に対応するセグメントに未処理のセグメントが存在するということであり、ステップ407へ進む。L\_seg\_cntがL\_seg\_num以上であれば、現在処理している単語に対応する全てのセグメントに対し処理を終えたということであり、ステップ410へ進む。

【0030】407では、表示すべき折れ線グラフの標本点を取得する。すなわち、L\_x[L\_seg\_cnt + 1]に、(L\_x[L\_seg\_cnt] + (音声データのセグメントの発話時間の配列のL\_seg\_cnt番目の要素)を代入し、L\_y[L\_seg\_cnt + 1]に、(音声データのセグメントのピッチ周波数の平均の配列のL\_seg\_cnt番目の要素)または、(音声データのセグメントの音圧の平均の配列のL\_seg\_cnt番目の要素)を代入する。イントネーションの折れ線グラフ表示時は、セグメントのピッチ周波数の平均の配列の要素、ストレスの折れ線グラフ表示時には、セグメントの音圧の平均の配列の要素を代入する。

【0031】408では、実際に折れ線グラフの表示を行う。すなわち、{(L\_x[L\_seg\_cnt], L\_y[L\_seg\_cnt]), (L\_x[L\_seg\_cnt + 1], L\_y[L\_seg\_cnt + 1])}の2点を結ぶ線分を表示する。

【0032】409では、X軸の下部に対応する単語文字列を表示するかどうかの判断を行う。つまり、disp\_word\_flagがONであるかどうか調べ、ONであればステップ410へOFFであれば、ステップ412へ進む。

【0033】410では、(L\_x[L\_seg\_cnt + 1], 0)の直下に音声データの単語文字列の配列22のL\_word\_cnt番目の要素である文字列を表示する。

【0034】411では、1つの単語につき単語文字列の表示は1度でいいのでdisp\_word\_flagをOFFにする。

【0035】412では、処理を次のセグメントに移すため、L\_seg\_cntに1を加え、ステップ406へ戻る。

【0036】413では、処理を次の単語に移すため、L\_word\_cntに1を加え、ステップ403へ戻る。

【0037】次に、データの具体例を用いて上記の処理をより具体的に説明する。

【0038】学習者のマイク15を使用して音声を入力すると、音声認識手段31でスペクトル分析等の手法により音声認識処理を行う。ここでは、音声認識手段31の出力として、図2に示した音声データが得られたとする。

【0039】グラフィック表示部32では、この音声デ

ータを入力として以下の処理を行う。

【0040】まず、前処理として、X軸、Y軸の表示、L\_word\_cnt、L\_seg\_cnt、L\_x[0]、L\_y[0]にそれぞれ0を代入する(ステップ401)。続いて、L\_word\_numに音声データの単語数の値である4を代入する(ステップ402)。

【0041】ここで、L\_word\_cntとL\_word\_numの比較を行う(ステップ403)が、今L\_word\_cntが0、L\_word\_numが4なので、ステップ404へ進む。

【0042】ステップ404では、今から処理する単語の文字列を次の折れ線表示時にX軸の下部に表示することを示すdisp\_word\_flagをONにする。次に、L\_seg\_numに音声データの各単語のセグメント数の配列のL\_word\_cnt番目の要素を加える。ここで、L\_seg\_num=0、L\_word\_cnt=0である。音声データの各単語のセグメント数の配列の0番目の要素は3であるのでL\_seg\_num=3となる(ステップ405)。

【0043】ステップ406では、L\_seg\_cntとL\_seg\_numの比較を行う。今、L\_seg\_cnt=0、L\_seg\_num=3なので、処理はステップ407へ進む。

【0044】ここで、実際に折れ線グラフ表示を行うための標本点を取得する。すなわち、L\_x[1]にL\_x[0]の値と音声データの各セグメントの発話時間の配列の0番目の要素の値を加え、24を代入する。また、例えばイントネーションのグラフを表示するのであれば、L\_y[1]に音声データの各セグメントのピッチ周波数の平均の配列の0番目の要素の値104を代入する(ステップ407)。続いて実際に、折れ線グラフを表示する。つまり、{(L\_x[0], L\_y[0]), (L\_x[1], L\_y[1])}の値である{(0, 0), (24, 104)}の2点を結ぶ線分を表示する(ステップ408)。

【0045】次に、disp\_word\_flagがONであるかどうかを調べ(ステップ409)、今ONなので、X軸の下部に対応する単語文字列を表示するための処理を行う。つまり、音声データの単語文字列の配列の0番目の値である“how”を(x[1], 0)の直下に表示する(ステップ410)。“how”を処理している間、もうこの文字列を表示する必要はないので、disp\_word\_flagをOFFにする(ステップ411)。

【0046】0番目のセグメントに対する処理が終了したので、L\_seg\_cntに1を加え1として、ステップ406へ戻る。1番目のセグメントについても上記と同様に折れ線グラフ表示を行う(ステップ407、408)。

【0047】ステップ409では、disp\_word\_flagがOFFになっているので、単語文字列の表示処理は行わず、ステップ412へ進む。

【0048】このようにして、2番目のセグメントまでの処理が終わったとする。このとき、L\_seg\_cntが3となっている(ステップ412)。ステップ406での比較の結果、L\_seg\_cntがL\_seg\_num以上になったので、ス

ステップ413へ進み、次の単語へ処理を進めるために、L\_word\_cntに1を加え、ステップ403へ戻る。

【0049】同様に、4番目の単語まで全ての処理が終わったとする。すると、L\_word\_cntが4となり（ステップ413）、403での比較の結果全ての単語に付いて処理が終わったとして、本実施例の処理が終了する。

【0050】（実施例2）実施例2は、第2の本発明の音声認識装置の一実施例であって、外国語発音練習において、手本となるある文章を入力した音声のイントネーションまたはストレスを、認識した単語に対応させてグラフィック表示し、さらに学習者の同じ文章の音声のイントネーションまたはストレスを、認識した単語を一致させ正規化して、単語毎に、手本のグラフィックに対応させて表示する装置の例である。

【0051】図7は、その実施例2のブロック図である。

【0052】音声認識手段51は、学習者が音声を入力し、認識結果として、音声データを出力する手段であり、実施例1の音声認識手段31と同一の機能を持つ。また、音声認識処理の結果得られる単語列は、無音部分を除くと、学習者により選択された練習文の練習文文字列61（図8参照）と同じであるとする。

【0053】練習文提示部52は、教材データ55に保存されている教材ファイルの練習文番号60と練習文文字列61を利用して練習文のリストを表示し、学習者にこれから練習する文と自分の性別の入力を促す。学習者により練習文が選択され、性別が入力されると、選択された練習文に対応する練習文番号と性別を出力する。

【0054】手本データ取得部53は、練習文提示部52の出力である学習者が選択した練習文の番号と学習者の性別を入力とし、教材データ55を読み込み、練習文の番号に対応する音声データファイルを手本音声データ56から得る手段である。

【0055】本発明の表示手段の一例としての正規化部およびグラフィック表示部54は、音声認識手段51から出力される学習者の音声データと手本データ取得部53からの手本音声データを入力とし、両者の対応する単語毎の発話時間を正規化し、学習者の音声データと手本の音声データ（イントネーション、ストレス）をグラフィック表示する手段である。図15及び図16は、そのようなグラフィック表示手段54の画面の一例である。

【0056】図8は、教材ファイルの例であり、練習文番号61、練習文の文字列62、男性用の音声データファイル名63、女性用の音声データファイル名64からなる。

【0057】図9は、手本音声データファイルの例であり、学習者の音声を認識した結果の音声データと同じ構造の手本音声データが保存されている。これら手本のデータは、予め第1の本発明で述べたようにして得られた

ものである。

【0058】図10は、正規化及びグラフィック表示部54の処理フローの例である。

【0059】801は、前処理であり、X軸、Y軸の表示、現在処理している学習者の音声データの単語が何番目かを示すL\_word\_cnt、現在処理している学習者の音声データのセグメントが何番目かを示すL\_seg\_cnt、学習者のイントネーション（ストレス）の折れ線グラフ表示時の標本点となる配列の0番目の要素L\_x[0]、L\_y[0]、現在処理している手本音声データの単語が何番目かを示すM\_word\_cnt、現在処理している手本音声データのセグメントが何番目かを示すM\_seg\_cnt、手本のイントネーション（ストレス）の折れ線グラフ表示時の標本点となる配列の0番目の要素M\_x[0]、M\_y[0]のそれぞれに0を代入する。

【0060】8011では、音声認識結果の音声データの単語数をL\_word\_numに代入する。

【0061】802は、全ての単語について、処理が終わったかどうかを調べるために、L\_word\_cntとL\_word\_numの比較を行う。L\_word\_cntがL\_word\_numより小さいとは、まだ処理すべき単語が残っているということであり、ステップ803へ進み、そうでないときは、全ての処理を終了する。

【0062】803では、手本音声データのM\_word\_cnt番目の単語について折れ線グラフと対応する単語文字列の表示を行う。このステップの詳細なフローの例を図11及び図12に示す。

【0063】805では、L\_word\_cnt番目の学習者の単語に対応するイントネーションまたはストレスを手本のイントネーションまたはストレスに対応させて表示するための倍率を求める。この処理の詳細フローの例を図13に示す。

【0064】806では、L\_word\_cnt番目の学習者の単語に対応するイントネーションまたはストレスを折れ線グラフ表示する。この時805で求めた倍率により手本の対応する単語と同じ位置（X座標）に表示する。

【0065】807では、学習者、手本共に次の単語へ処理を進めるため、L\_word\_cnt、M\_word\_cntの値にそれぞれ1を加える。

【0066】図11及び図12は、図10のステップ803の詳細フローの例である。

【0067】901では、次の折れ線グラフ表示時にX軸の下部に単語文字列を表示することを示すdisp\_word\_flagをONにする。また、この単語を発話するのに要した時間を示す変数M\_word\_durに0を代入する。

【0068】902では、現在処理している単語が何番目のセグメントまでかを示す変数M\_seg\_numに音声データの各単語のセグメント数の配列のM\_word\_cnt番目の要素を加える。

【0069】903では、M\_seg\_cntとM\_seg\_numとの比

較を行う。 $M\_seg\_cnt$ が $M\_seg\_num$ より小さければ、現在処理している単語に対応するセグメントに未処理のセグメントが存在するということであり、ステップ904へ進む。 $M\_seg\_cnt$ が $M\_seg\_num$ 以上であれば、現在処理している単語に対応する全てのセグメントに対し処理を終えたということであり、処理を終了する。

【0070】904では、表示すべき折れ線グラフの標本点を取得する。すなわち、 $M\_x[M\_seg\_cnt + 1]$ に、 $\{M\_x[M\_seg\_cnt] + (\text{音声データのセグメントの発話時間の配列の}M\_seg\_cnt\text{番目の要素})\}$ を代入し、 $M\_y[M\_seg\_cnt + 1]$ に、 $(\text{音声データのセグメントのピッチ周波数の平均の配列の}M\_seg\_cnt\text{番目の要素})$ または、 $(\text{音声データのセグメントの音圧の平均の配列の}M\_seg\_cnt\text{番目の要素})$ を代入する。イントネーションの折れ線グラフ表示時は、セグメントのピッチ周波数の平均の配列の要素、ストレスの折れ線グラフ表示時には、セグメントの音圧の平均の配列の要素を代入する。

【0071】また、同時に $M\_word\_dur$ に音声データのセグメントの発話時間の配列の $M\_seg\_cnt$ 番目の要素を加える。

【0072】905では、実際に折れ線グラフの表示を行う。すなわち、 $\{(M\_x[M\_seg\_cnt], M\_y[M\_seg\_cnt]), (M\_x[M\_seg\_cnt + 1], M\_y[M\_seg\_cnt + 1])\}$ の2点を結ぶ線分を表示する。

【0073】906では、X軸の下部に対応する単語文字列を表示するかどうかの判断を行う。つまり、 $disp\_word\_flag$ がONであるかどうか調べ、ONであればステップ907へOFFであれば、ステップ909へ進む。

【0074】907では、 $(M\_x[M\_seg\_cnt], 0)$ の直下に認識結果の単語文字列の配列22の $M\_word\_cnt$ 番目の要素である文字列を表示する。

【0075】908では、1つの単語につき単語文字列の表示は1度でいいので $disp\_word\_flag$ をOFFにする。

【0076】909では、処理を次のセグメントに移すため、 $M\_seg\_cnt$ に1を加え、ステップ903へ戻る。

【0077】図13は、図10のステップ805の詳細フローの例である。

【0078】1001では、現在処理している学習者の単語のセグメント数を示す変数 $L\_word\_seg\_num$ に音声データの各単語のセグメント数の配列の $L\_word\_cnt$ 番目の値を代入し、また、この単語を発話するのに要した時間を示す変数 $L\_word\_dur$ 、ループ用の変数 $i$ をそれぞれ0に初期化する。

【0079】1002では、 $i$ と $L\_word\_seg\_num$ の比較を行う。 $i$ が $L\_word\_seg\_num$ より小さければステップ1003へ進む、 $i$ が $L\_word\_seg\_num$ 以上であればステップ1005へ進む。

【0080】1003では、現在処理している単語の $i$ 番目のセグメントの時間を $L\_word\_dur$ に加える。すなわ

ち、学習者の音声データの各セグメントの発話時間の配列の $(L\_seg\_cnt + i)$ 番目の値を $L\_word\_dur$ に加える。

【0081】1004では、次のセグメントの発話時間を求めるため $i$ に1を加える。

【0082】1005では、現在処理している学習者の単語を表示するときの倍率を求め、処理を終了する。つまり、倍率を示す変数 $rate$ に $(M\_word\_dur / L\_word\_dur)$ を代入し、処理を終える。

【0083】図14は、図10のステップ806の詳細フローの例である。

【0084】1101では、現在処理中の単語に対応するセグメントが全体で何番目のセグメントまでかを示す変数 $L\_seg\_num$ に音声データの各単語のセグメント数の配列の $L\_word\_cnt$ 番目の要素の値を加える。

【0085】1102では、現在処理を行っている単語に対応する全てのセグメントに対し処理を終えたかどうかを調べる。すなわち、 $L\_seg\_cnt$ と $L\_seg\_num$ の値を比較し、 $L\_seg\_cnt$ が $L\_seg\_num$ より小さければステップ1103へ進む、そうでなければ処理を終了する。

【0086】1103では、表示すべき折れ線グラフの標本点を取得する。すなわち、 $L\_x[L\_seg\_cnt + 1]$ に、 $\{L\_x[L\_seg\_cnt] + (\text{音声データのセグメントの発話時間の配列の}L\_seg\_cnt\text{番目の要素}) * rate\}$ を代入し、 $L\_y[L\_seg\_cnt + 1]$ に、 $(\text{音声データのセグメントのピッチ周波数の平均の配列の}L\_seg\_cnt\text{番目の要素})$ または、 $(\text{音声データのセグメントの音圧の平均の配列の}L\_seg\_cnt\text{番目の要素})$ を代入する。イントネーションの折れ線グラフ表示時は、セグメントのピッチ周波数の平均の配列の要素、ストレスの折れ線グラフ表示時には、セグメントの音圧の平均の配列の要素を代入する。

【0087】1104では、実際に折れ線グラフの表示を行う。すなわち、 $\{(L\_x[L\_seg\_cnt], L\_y[L\_seg\_cnt]), (L\_x[L\_seg\_cnt + 1], L\_y[L\_seg\_cnt + 1])\}$ の2点を結ぶ線分を表示する。

【0088】1105では、処理を次のセグメントに移すため、 $M\_seg\_cnt$ に1を加え、ステップ1102へ戻る。

【0089】次に、データ的具体例を用いて上記の処理を説明する。

【0090】練習文提示部52により、図8の教材ファイルをもとに練習文リストが学習者に提示されたとする。

【0091】続いて、学習者により、練習文として、“how do you do”が選択され、また、学習者は男性であるとする。

【0092】手本データ取得部53は、教材ファイル55を参照し、“how do you do”の男性用の手本音声データファイル名“M\_ml.dat”を得る。“M\_ml.dat”の内容は、図9に示す手本データファイルの例と同一であるとする。

【0093】さらに、学習者の音声入力による音声認識手段51の出力は、図2に示すとおりであるとする。

【0094】正規化及びグラフィック表示部54では、学習者音声データ、手本音声データを入力とし、以下のように処理を行う。

【0095】まず、X軸、Y軸の表示、各種変数の初期化を行う(ステップ801)。

【0096】次に、L\_word\_numに学習者の音声データの単語数を代入する。ここでは、4が代入される(ステップ8011)。

【0097】ここで、L\_word\_cntとL\_word\_numの比較を行う(ステップ802)が、今L\_word\_cntが0、L\_word\_numが4なので、ステップ803へ進む。

【0098】ステップ803では、L\_word\_cnt番目の手本音声データの単語に対応するイントネーション(ストレス)を表示する。

【0099】すなわち、ステップ901では、手本におけるその単語の発話時間を示す変数M\_word\_durを0に初期化し、さらに今から処理する単語の文字列を次の折れ線グラフ表示時にX軸の下部に表示することを示すdisp\_word\_flagをONにする。

【0100】次に、M\_seg\_numに音声データの各単語のセグメント数の配列のL\_word\_cnt番目の要素を加える。ここで、M\_seg\_num=0、L\_word\_cnt=0である。音声データの各単語のセグメント数の配列の0番目の要素は2であるのでM\_seg\_num=2となる(ステップ902)。

【0101】ステップ903では、M\_seg\_cntとM\_seg\_numの比較を行う。今、M\_seg\_cnt=0、M\_seg\_num=2なので、処理はステップ904へ進む。

【0102】ここで、実際に折れ線グラフ表示を行うための標本点を取得する。すなわち、M\_x[1]にM\_x[0]の値と音声データの各セグメントの発話時間の配列のM\_seg\_cnt=0番目の要素の値を加え、14を代入する。また、例えばイントネーションのグラフを表示するのであれば、M\_y[1]に音声データの各セグメントのピッチ周波数の平均の配列の0番目の要素の値114を代入する。さらに、M\_word\_durに現在のM\_word\_durに手本音声データの各セグメントの発話時間の配列のM\_seg\_cnt=0番目の値14を加え、14とする。(ステップ904)。

【0103】続いて実際に、折れ線グラフを表示する。つまり、{(L\_x[0], L\_y[0]), (L\_x[1], L\_y[1])}の値である{(0, 0), (14, 114)}の2点を結ぶ線分を表示する(ステップ905)。

【0104】次に、disp\_word\_flagがONであるかどうかを調べ(ステップ906)、今ONなので、X軸の下部に対応する単語文字列を表示するための処理を行う。つまり、音声データの単語文字列の配列の0番目の値である"how"を(x[1], 0)の直下に表示する(ステップ907)。

"how"を処理している間、もうこの文字列を表示

する必要はないので、disp\_word\_flagをOFFにする(ステップ908)。

【0105】0番目のセグメントに対する処理が終了したので、M\_seg\_cntに1を加え1として(ステップ909)、ステップ903へ戻る。1番目のセグメントについても上記と同様に折れ線グラフ表示を行う(ステップ905、906)。

【0106】ステップ409では、disp\_word\_flagがOFFになっているので、単語文字列の表示処理は行わず、ステップ909へ進む。

【0107】このとき、M\_seg\_cntが2となっている(ステップ909)。ステップ903での比較の結果、M\_seg\_cntがM\_seg\_num以上になったので、手本の0番目の単語に対する処理を終え、ステップ805へ進む。ここで、M\_word\_dur=27となっている。

【0108】次に、学習者のL\_word\_cnt番目の単語に対応するセグメントのデータを手本に対応させて表示するための倍率を計算する(ステップ805)。

【0109】すなわち、学習者のL\_word\_cnt番目の単語に対応するセグメント数を示す変数L\_word\_seg\_numに音声データの各単語のセグメント数の配列のL\_word\_cnt=0番目の値3を代入し、L\_word\_dur、iにそれぞれ0を代入する。

【0110】次に、L\_word\_cnt=0番目の単語に対応する全てのセグメントに対して処理を終えたかどうか調べるために、iとL\_word\_seg\_numを比較する。今、i=0、L\_word\_seg\_num=3なので、処理はステップ1003へ進む。

【0111】ここで、L\_word\_durに現在のL\_word\_dur=0と音声データの各セグメントの発話時間の配列の(L\_seg\_cnt + i)=0番目の値24の和である24を代入する(ステップ1003)。

【0112】さらに、次のセグメントの発話時間を調べるためiに1を加える(ステップ1004)。

【0113】このようにして、処理を2番目のセグメントまで終えたとする。このとき、i=3、L\_word\_dur=46となり、ステップ1002の比較により、ステップ1005へ進む。

【0114】ステップ1005では、(M\_word\_dur=27)/(L\_word\_dur=46)をrateに代入する。すなわち、rateに約0.59が代入される。

【0115】学習者の音声データを表示する際の倍率が求められた(ステップ805)ので、この倍率を利用して、L\_word\_cnt=0番目の学習者の単語に対応するイントネーション(ストレス)のグラフィック表示を行う(ステップ806)。

【0116】ステップ1101では、L\_seg\_numに音声データの各単語のセグメント数の配列のL\_word\_cnt番目の要素を加える。ここで、L\_seg\_num=0、L\_word\_cnt=0である。音声データの各単語のセグメント数の配列

の0番目の要素は3であるのでL\_seg\_num=3となる。

【0117】ステップ1102では、L\_seg\_cntとL\_seg\_numの比較を行う。今、L\_seg\_cnt=0、L\_seg\_num=3なので、処理はステップ1103へ進む。

【0118】ここで、実際に折れ線グラフ表示を行うための標本点を取得する。すなわち、L\_x[1]にL\_x[0]の値に音声データの各セグメントの発話時間の配列の0番目の要素の値と倍率rateの積を加える。ここでは、 $0+24 \times 0.59=14.16$ を代入する。また、例えばイントネーションのグラフを表示するのであれば、L\_y[1]に音声データの各セグメントのピッチ周波数の平均の配列の0番目の要素の値104を代入する(ステップ1103)。続いて実際に、折れ線グラフを表示する。つまり、{(L\_x[0], L\_y[0]), (L\_x[1], L\_y[1])}の値である{(0, 0), (14.16, 104)}の2点を結ぶ線分を表示する(ステップ1104)。

【0119】学習者のL\_seg\_cnt=0番目のセグメントのデータの表示が終わったので、処理を次のセグメントに進めるため、L\_seg\_cntに1を加える(ステップ1105)。

【0120】このようにして、処理を進め、学習者の2番目のセグメントまでの処理が終わるとL\_seg\_cnt=3となり、ステップ1102の比較の結果、学習者のL\_word\_cnt=0番目の単語の処理を終わり、ステップ807へ進む。

【0121】ステップ807では、処理を次の単語に進めるため、L\_word\_cntに1を加え、1とし、ステップ802へ戻る。

【0122】このようにして処理を進め、3番目の単語まで処理を終えたとき、L\_word\_cnt=4となり、ステップ802の比較により、本実施例の全ての処理を終了する。

【0123】なお、第3の本発明として、手本となる人の音響的特徴量と言語シンボルとの対応付け内容と、学習者のそれとを、上記実施例のように、正規化して表示するのではなく、あるいは表示に加えて、両者の対応付け内容を比較手段で比較して、その不一致の部分について、ある単語のイントネーションを下げるようになどという指示を表示するようにしてもよい。図16の130はその指示の一例である。

【0124】また、手本となる音声データなどは、予め記憶手段に格納されてなく、いつでもネイティブスピーカーなどによって入力でき、音声信号の所定の音響的特徴量とそれに対応する言語シンボルとの対応付け内容が得られるようになっていてもよい。

【0125】また、本発明の音響的特徴量とは、イントネーションやストレスに限らず、フォルマント等の他の音響的特徴量であってもよい。

【0126】また、本発明の言語的シンボルとは、言葉、発音記号、音素コード、音素片など、言語に関する

シンボルであればどのようなものでもよい。

【0127】また、本発明の各手段は、コンピュータを用いてソフトウェア的に実現しても、それら機能を有する専用のハード回路を用いて実現してもかまわない。

【0128】

【発明の効果】以上の説明から明らかなように、第1の本発明は、音声信号の所定の音響的特徴量とそれに対応する言語シンボルとを対応付けながら表示する対応表示手段を備えるので、その音響的特徴量と言語シンボルとの対応関係が分かりやすいという長所を有する。

【0129】また、第2の本発明は、音声信号の所定の音響的特徴量とそれに対応する言語シンボルとを対応付ける対応手段と、少なくとも2種類の音声について、対応手段によって得られた対応付け内容を比較する比較手段とを備えるので、例えば、1種類の音声をネイティブスピーカーの音声とし、他方の種類の音声を学習者の音声とすると、それらの比較によって、学習者の悪い所などを的確に指摘することが可能になる。また、第3の本発明は、音声信号の所定の音響的特徴量とそれに対応する言語シンボルとを対応付ける対応手段と、少なくとも2種類の音声について、各対応手段によって得られた対応付け内容を正規化して表示する表示手段とを備えるので、例えば、1種類の音声をネイティブスピーカーの音声とし、他方の種類の音声を学習者の音声とすると、それらの対応付け内容が正規化されて表示されるので、その表示を見て、学習者の悪い所などを的確に理解することが可能になる。

【図面の簡単な説明】

【図1】本発明の音声認識装置の一実施例を示すブロック図である。

【図2】本発明の音声認識手段の音声データである。

【図3】第1の本発明の音声認識装置の一実施例を示すブロック図である。

【図4】本発明の音声認識手段の動作を示すフローチャートである。

【図5】第1の本発明の対応表示手段の動作を説明するためのフローチャートである。

【図6】第1の本発明の対応表示手段の動作を説明するためのフローチャートである。

【図7】第2の本発明の音声認識装置の一実施例を示すブロック図である。

【図8】同実施例の教材ファイルの一例を示す構成図である。

【図9】同実施例の手本音声データファイルの一例を示す構成図である。

【図10】同実施例の動作を説明するためのフローチャートである。

【図11】図10のステップ803の詳細フローチャートである。

【図12】図10のステップ803の詳細フローチャー

トである。

【図13】図10のステップ805の詳細フローチャートである。

【図14】図10のステップ806の詳細フローチャートである。

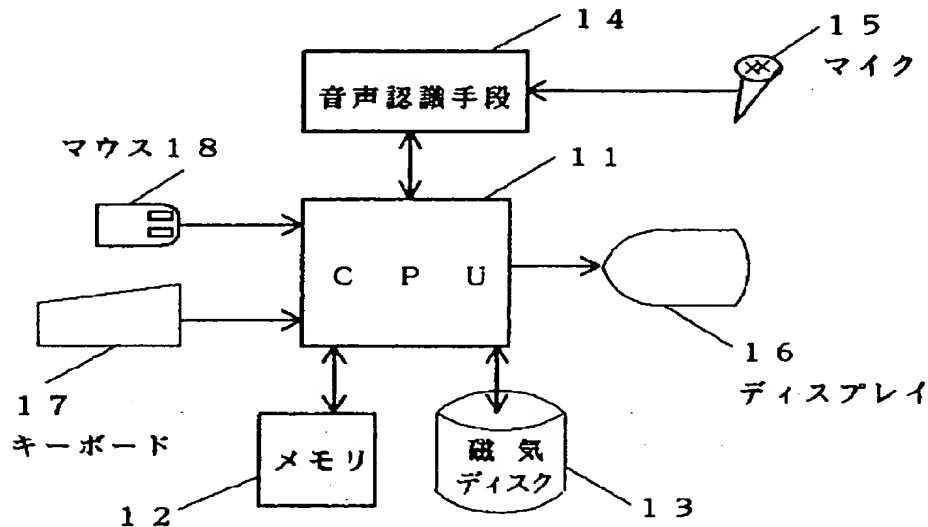
【図15】第2の本発明の音声認識装置の表示画面を示す図である。

【図16】第2の本発明の音声認識装置の表示画面を示す図である。

【符号の説明】

30	入力手段
31	音声認識手段
32	グラフィック表示部
50	入力手段
51	音声認識手段
52	練習文提示部
53	手本データ取得部
54	正規化及びグラフィック表示部
55	教材データ
10 56	手本音声データ

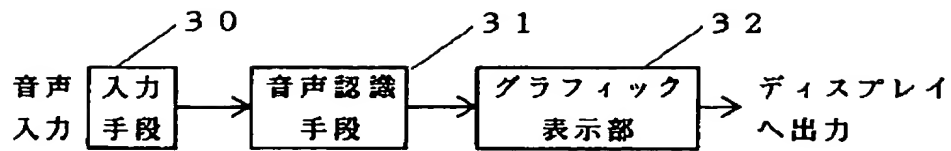
【図1】



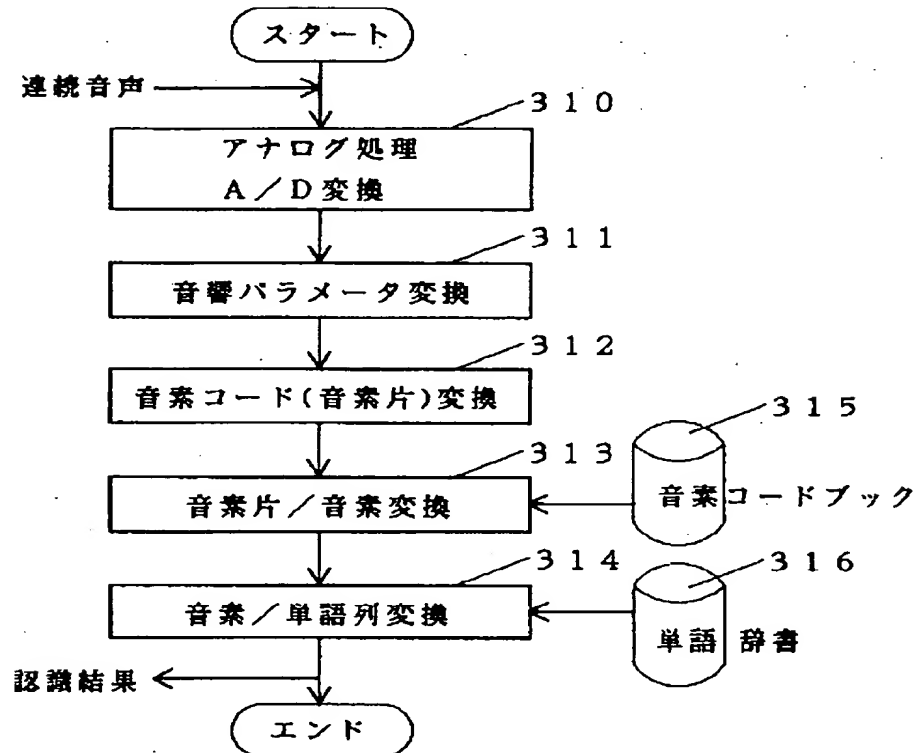
【図2】

単語数	21	4
単語文字列	22	[how][do][you][do]
各単語のセグメント数	23	[3][2][2][3]
各セグメントのピッチ	24	[104][100][102][110][130][100][100][106][200][150]
各セグメントの音圧	25	[154][200][12][140][150][140][200][126][230][57]
各セグメントの発話時間	26	[24][10][12][11][13][10][10][18][20][15]

【図3】



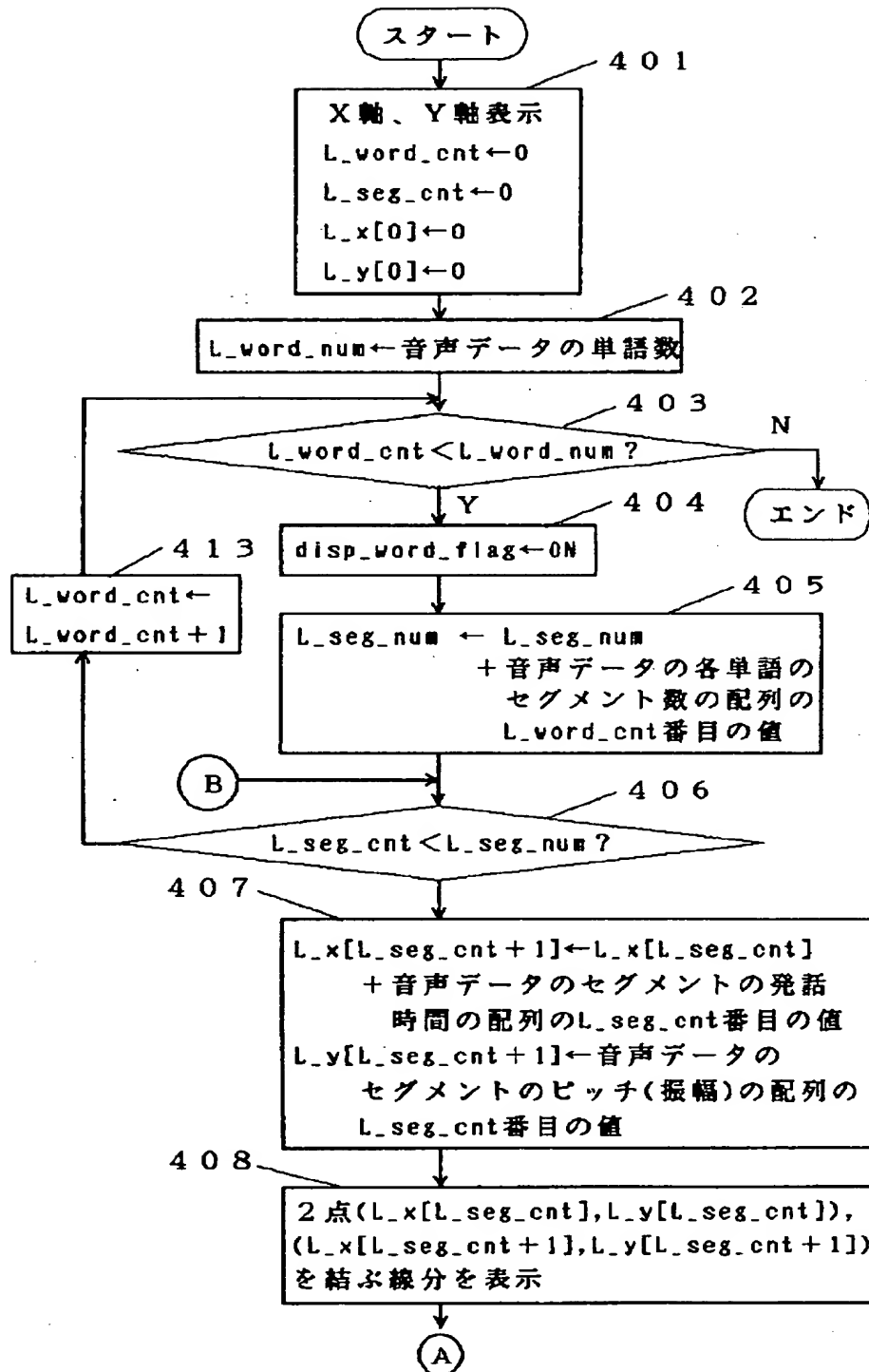
【図4】



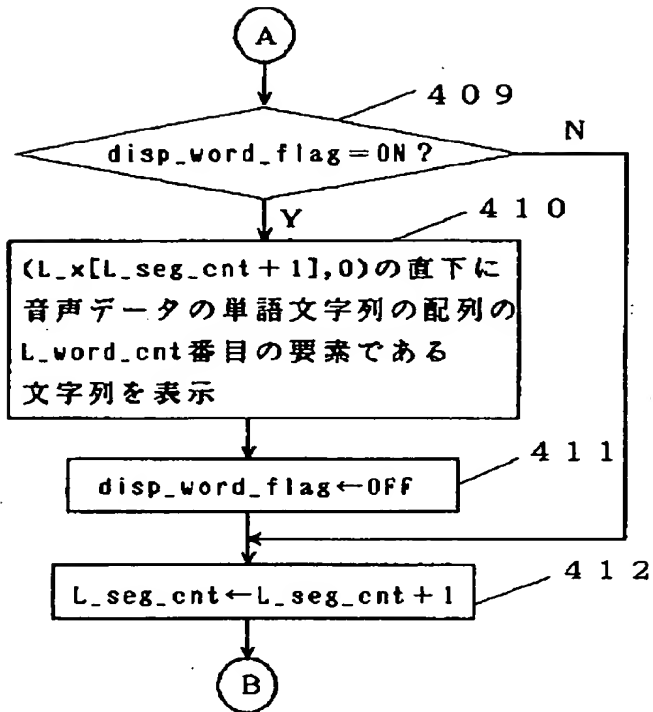
【図8】

練習文 番号	練習文文字列	男性用手本音声 データファイル名	女性用手本音声 データファイル名
1	"how do you do"	"M_m1.dat"	"M_f1.dat"
2	"good morning"	"M_m2.dat"	"M_f2.dat"
3	"good bye"	"M_m3.dat"	"M_f3.dat"

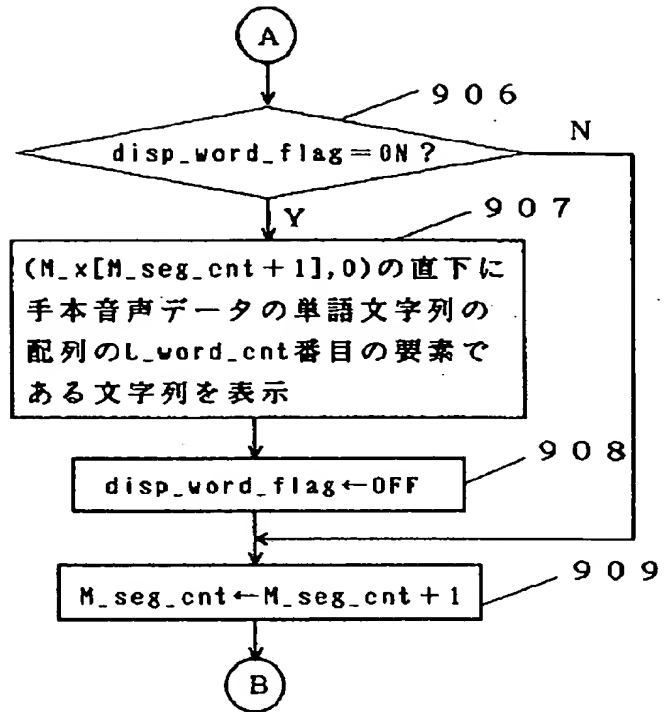
【図5】



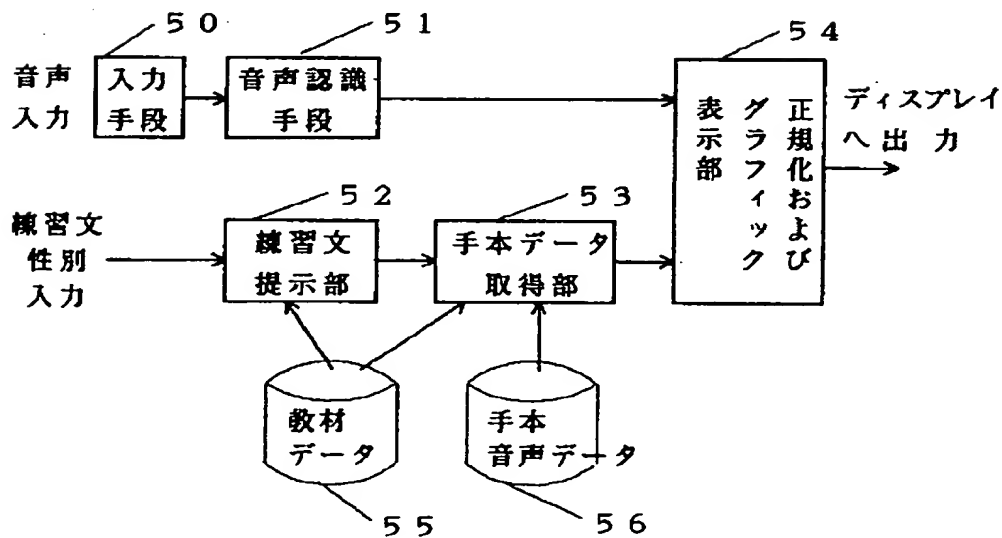
【図6】



【図12】



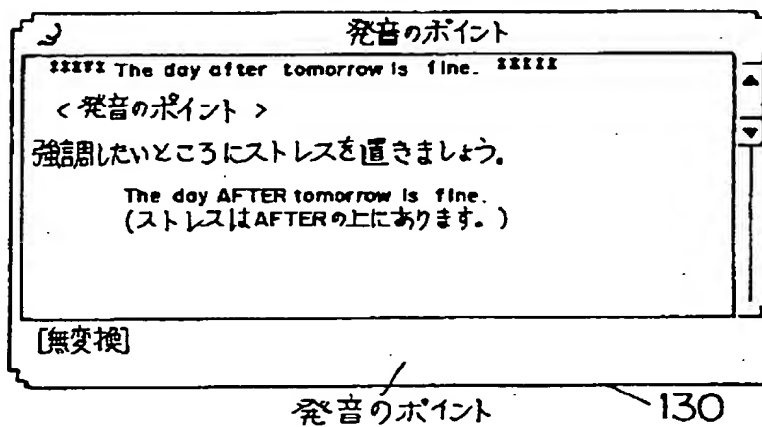
【図7】



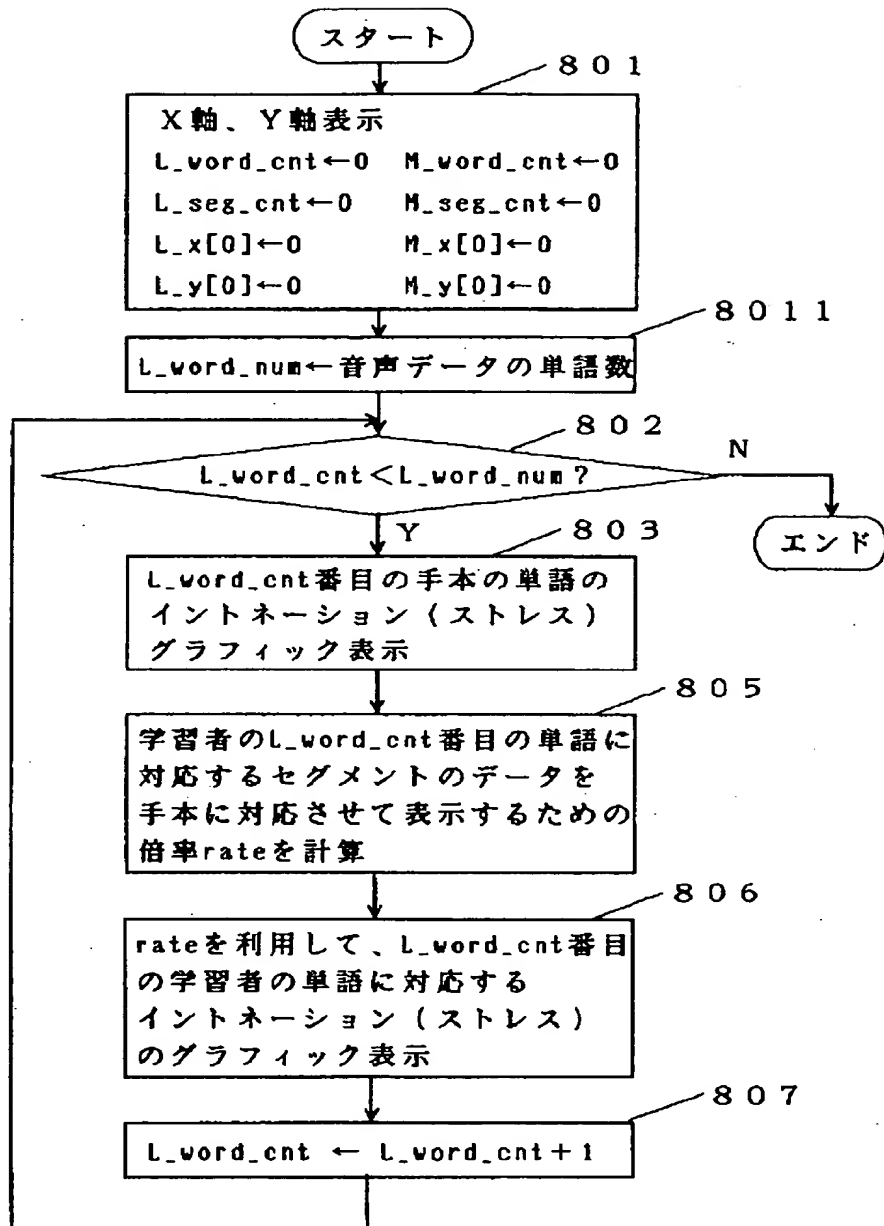
【図9】

単語数	7 1	4	
単語文字列	7 2	[how][do][you][do]	
各単語の セグメント数	7 3	[ 2 ][ 1 ][ 2 ][ 3 ]	
各セグメント のピッチ	7 4	[114][100]	[100][100][ 98][108][200][150]
各セグメント の音圧	7 5	[144][100]	[110][208][158][ 86][130][ 97]
各セグメント の発話時間	7 6	[ 14][ 13]	[ 10][ 10][ 9][ 15][ 13][ 14]

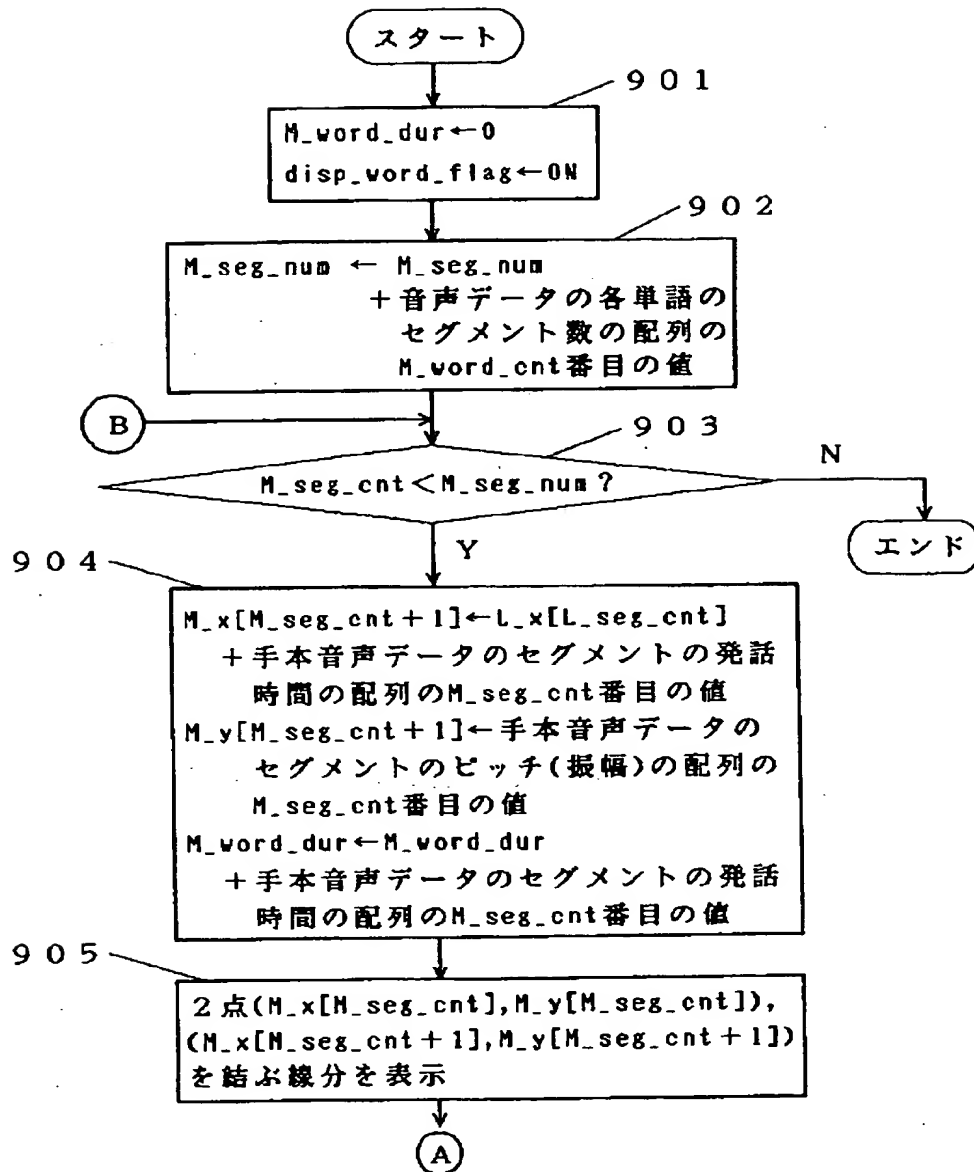
【図16】



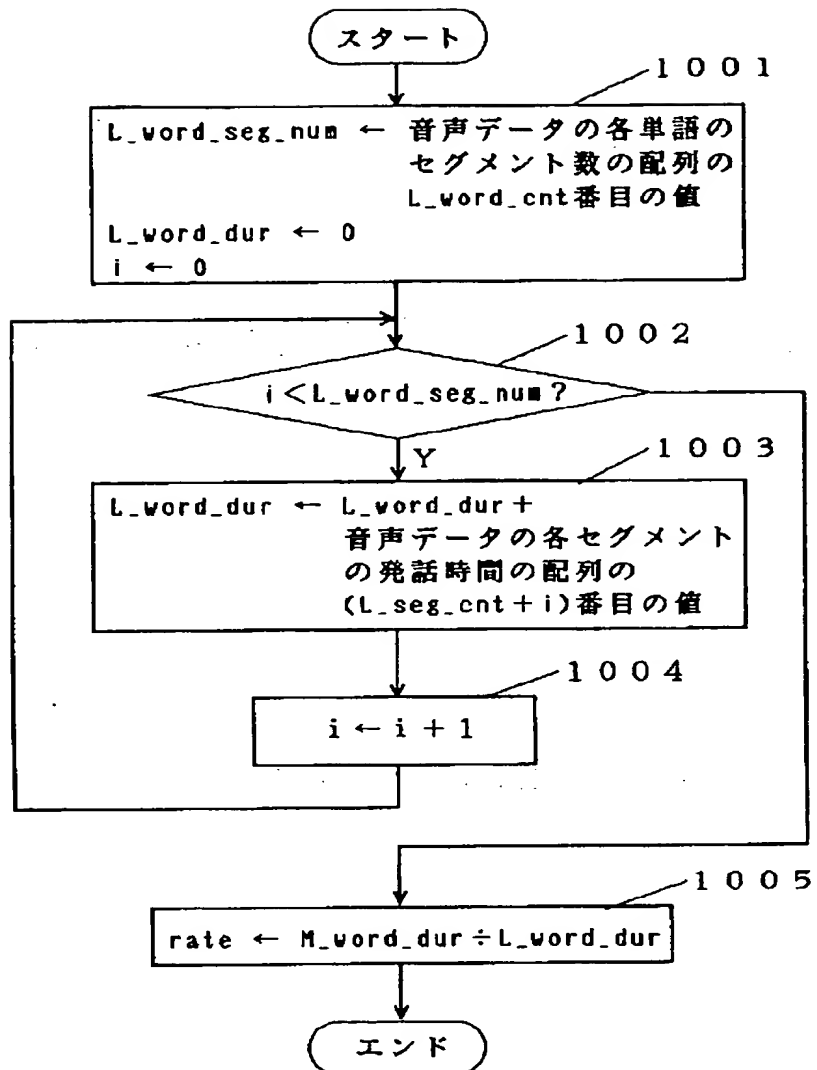
【図10】



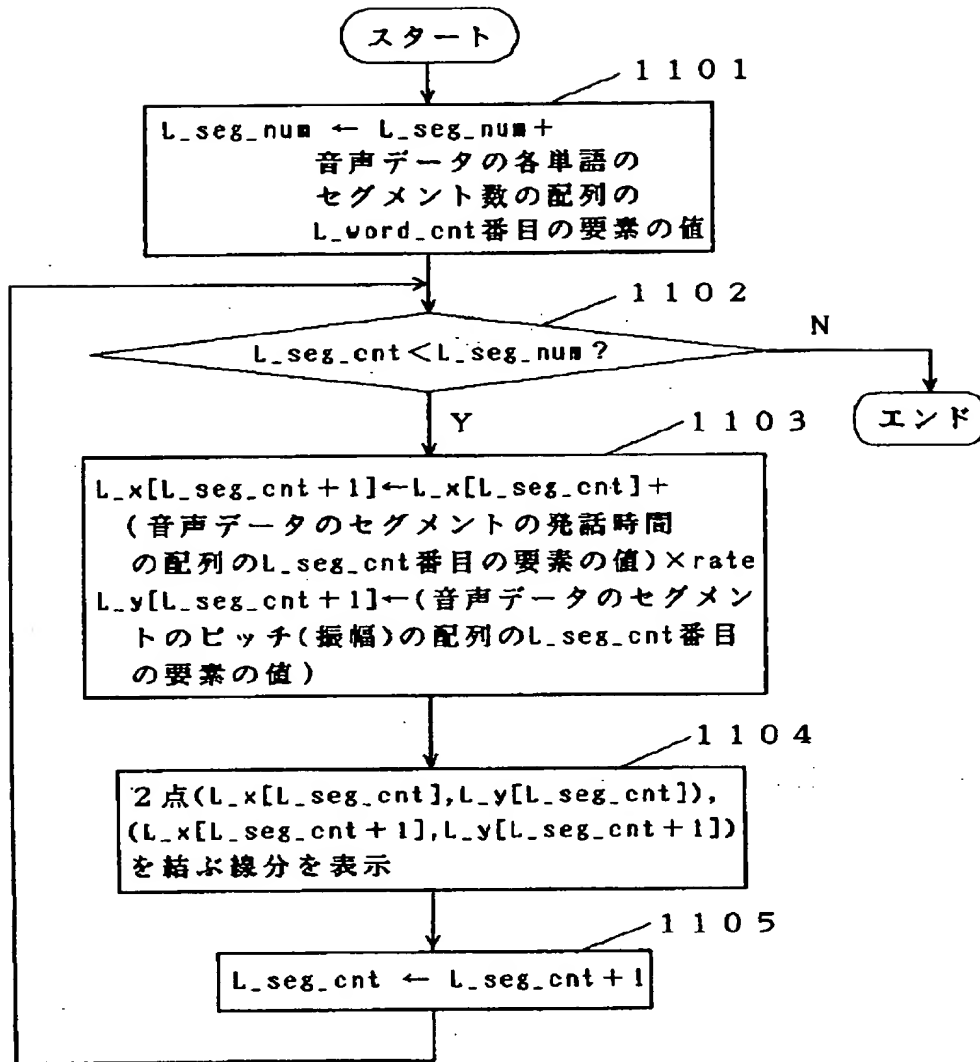
【図11】



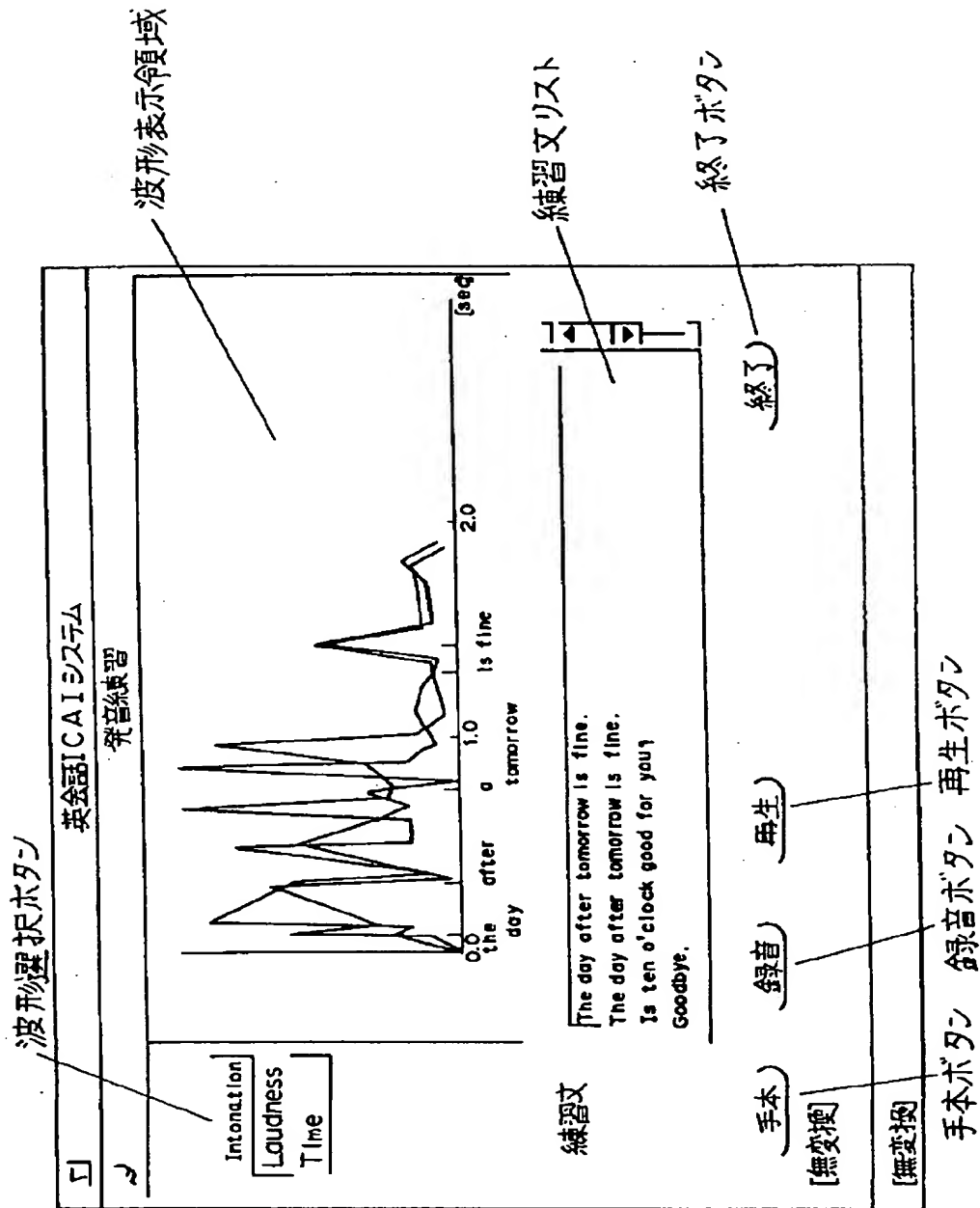
【図13】



【図14】



【図15】



フロントページの続き

(72)発明者 田川 忠道  
東京都港区虎ノ門1丁目7番12号 沖電気  
工業株式会社内

(72)発明者 加藤 正明  
愛知県名古屋市千種区内山三丁目8番10号  
株式会社沖テクノシステムズラボラトリ  
内